Clinical potential of whole-genome data linked to mortality statistics in patients with breast cancer in the UK: a retrospective analysis

Daniella Black, Helen Ruth Davies, Gene Ching Chiek Koh, Lucia Chmelova, Marko Cubric, Georgia Chalivelaki Chan, Andrea Degasperi, Jan Czarnecki, Ping Jing Toong, Yasin Memari, James Whitworth, Salome Jingchen Zhao, Yogesh Kumar, Shadi Basyuni, Giuseppe Rinaldi, Scott Shooter, Vladyslav Dembrovskyi, Rosie Davies, Maria Chatzou Dunford, Ellen Copson, Carlo Palmieri, Åke Borg, John Ambrose, Catey Bunce, Alona Sosinsky, Prabhu Arumugam, Matthew Arthur Brown, Johan Staaf, Nicholas Turner, Serena Nik-Zainal

Summary

Background Breast cancer is the most frequently diagnosed cancer in women. Survival is generally considered favourable, yet some patients remain at risk of early death. We aimed to assess whether comprehensive whole-genome sequencing (WGS) linked to mortality data could add prognostic value to existing clinical measures and identify patients who might respond to targeted therapeutics.

Methods In this integrative, retrospective analysis, we analysed 2445 breast cancer tumours (any stage and molecular subtype) collected from 2403 patients recruited through 13 National Health Service Genomic Medicine Centres or hospitals in England affiliated to the 100 000 Genomes Project (100kGP) between 2012 and 2018. We linked 2208 (90%) cases with clinical data; mortality data were obtained for 1188 patients. Following high-depth WGS of tumour and matched normal DNA, we performed comprehensive WGS profiling seeking driver mutations, mutational signatures, and compound algorithmic scores for homologous recombination repair deficiency (HRD), mismatch repair deficiency, and tumour mutational burden. Data from 1803 additional patients with breast cancer from three independent cohorts were used to validate various findings. To evaluate the prognostic value of WGS features, we performed univariable and multivariable Cox regression on data from patients with stage I–III, ER-positive, HER2-negative breast cancer with a cancer-specific mortality endpoint (around 5-year follow-up).

Findings Among 2445 tumours in the 100kGP breast cancer cohort, we observed genomic characteristics with immediate personalised medicine potential in 656 (26·8%), including features reporting HRD (298 [12·2%] total cases and 76 [6·3%] ER-positive, HER2-negative cases), highly individualised driver events, mutations underpinning resistance to endocrine therapy, and mutational signatures indicating therapeutic vulnerabilities. 373 (15·2%) cases had WGS features with potential for translational research, including compromised base excision repair and non-homologous end-joining dependency. Structural variation burden (hazard ratio 3·9 [95 CI% 2·4–6·2]; p<0·0001), high levels of APOBEC signatures (2·5 [1·6–4·1]; p<0·0001), and *TP53* drivers (3·9 [2·4–6·2]; p<0·0001) were independently prognostic of customary clinical measures (age at diagnosis, stage, and grade) in patients with ER-positive, HER2-negative breast cancer. We developed a prognosticator for ER-positive, HER2-negative breast cancer capable of identifying patients who require either increased intervention or therapy de-escalation, validating the framework in the independent Swedish Cancerome Analysis Network-Breast (SCAN-B) dataset.

Interpretation We show that breast cancer genomes are rich in predictive and prognostic value. We propose a two-step model for effective clinical application. First, the identification of candidates for targeted therapies or clinical trials using highly individualised genomic markers. Second, for patients without such features, the implementation of enhanced prognostication using genomic features alongside existing clinical decision-making factors.

Funding National Institute of Health Research, Breast Cancer Research Foundation, Dr Josef Steiner Cancer Research Award 2019, Basser Gray Prime Award 2020, Cancer Research UK, Sir Jeffrey Cheah Early Career Fellowship, the Mats Paulsson Foundation, the Fru Berta Kamprads Foundation, and the Swedish Research Council.

Introduction

Globally, there were 2·3 million women diagnosed with breast cancer and 670 000 related deaths in 2022.[1] Accurately selecting therapeutic strategies for individual patients and identifying those with likely poor prognosis remain challenging. Most patients present with early-stage breast cancer and treatment decisions are variably informed by clinical and histopathological characteristics, such as lymph node involvement; age at diagnosis; tumour grade, size, and stage; and ER and HER2 status.[2]

Lancet Oncol 2025; 26: 1417–31

Published **Online** October 7, 2025 https://doi.org/10.1016/S1470-2045(25)00400-0

See **Comment** page 1394

Department of Genomic Medicine, University of Cambridge, Cambridge, UK (D Black MSc, H R Davies PhD, G C C Koh PhD, L Chmelova MPhil, A Degasperi PhD, J Czarnecki PhD, P J Toong BA, Y Memari PhD, J Whitworth MD, S J Zhao MPhil, Y Kumar PhD, S Basyuni MD, G Rinaldi MSc, S Shooter MSc, Prof S Nik-Zainal MD PhD); **Early Cancer Institute, Department of Oncology, University of Cambridge, Cambridge, UK** (D Black MSc, H R Davies PhD, G C C Koh PhD, L Chmelova MPhil, A Degasperi PhD, J Czarnecki PhD, P J Toong BA, Y Memari PhD, S J Zhao MPhil, Y Kumar PhD, S Basyuni MD, G Rinaldi MSc, S Shooter MSc, Prof S Nik-Zainal MD PhD); **Sir Jeffrey Cheah Sunway Medical School, Faculty of Medical and Life Sciences, Sunway University, Sunway City, Malaysia** (G C C Koh PhD); **Genomics England, London, UK** (M Cubric MSc, G Chalivelaki Chan PhD, J Ambrose PhD, A Sosinksy PhD, P Arumugam MD PhD, M A Brown MD); **Lifebit Biotech, London, UK** (V Dembrovskyi MSc, R Davies PhD, M Chatzou Dunford PhD); **Cancer Sciences Academic Unit, University of Southampton, Southampton, UK** (Prof E Copson MD PhD); **The Clatterbridge Cancer Centre, Wirral, UK** (Prof C Palmieri MD PhD); **Department of Molecular and Clinical Cancer, Medicine,**

Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK (Prof C Palmieri MD PhD); **Division of Oncology & Pathology, Department of Clinical Sciences, Lund University, Lund, Sweden** (Prof Å Borg PhD, J Staaf PhD); **Breast Unit, Royal Marsden Hospital and Institute of Cancer Research, London, UK** (C Bunce DSc, Prof N Turner MD PhD); **NIHR Institute of Cancer Research BRC, London, UK** (C Bunce DSc, Prof N Turner MD PhD)

Correspondence to: Prof Serena Nik-Zainal, Department of Genomic Medicine, University of Cambridge, Cambridge CB2 0QQ, UK
sn206@cam.ac.uk

Research in context

Evidence before this study

Whole-genome sequencing (WGS) of breast cancers has helped to shed light on the totality of constitutional and somatic driver events and mutational processes that shape breast cancer. We searched PubMed for articles published from database inception to May 16, 2025, using the search terms: "breast cancer" AND "whole genome sequencing" AND "clinical trial". This search yielded only 12 results, suggesting that the scarcity of systematically linked data on clinical outcomes and mortality in most WGS studies and, crucially, the under-application of WGS in clinical trials, has limited the clinical utility offered by WGS to date. Thus, the prognostic or predictive value of WGS has remained low, which is particularly important in breast cancer research where clinical gaps remain. One such gap is understanding the reasons behind premature mortality among some patients with low-risk tumours by clinical metrics; another is predicting patient subgroups who are more likely to respond effectively to targeted therapeutics.

Added value of this study

To our knowledge, this integrative, retrospective analysis is the largest WGS study of a population-based cohort of patients with breast cancer (approximately 2500 patients). First, we present a simple WGS prognostic risk framework for patients with ER-positive, HER2-negative breast cancer, which adds value beyond the customary clinical markers currently used in the National Health Service (NHS) in the UK. Designed to be used alongside existing clinical metrics, this prognostic framework could have a considerable impact on several key patient groups: patients at low risk clinically and high risk genomically, who require additional intervention and monitoring; patients at low risk clinically and genomically, who can be given more certainty about their trajectory or might be potential candidates for treatment de-escalation; and patients at high risk clinically and genomically with no other targetable abnormality identified, warranting new therapeutic strategies. Second, we were able to comprehensively catalogue WGS markers with potential for precision medicine in 27% of tumours in this cohort. These highly individualised markers obtainable from a single readout could be used as a triage tool to predict response to targeted therapeutics, highlight patients at risk of treatment resistance, and inform recruitment to prospective clinical trials.

Implications of all the available evidence

WGS offers the ability to distinguish clinically important subsets of patients with breast cancer at a time when logistical challenges to the widespread implementation of WGS in the clinic have diminished. Because WGS offers an all-inclusive readout of genomic abnormalities in a single assay, our work requests a mindset shift in how to use genomic information for breast cancer in the clinic. We put forth a blueprint for using WGS as a triaging step in clinical care and for clinical trials in patients with breast cancer.

Genomics has begun to inform cancer management. However, critics contend that genomics has under-delivered on the promise of personalised medicine,[3,4] which is arguably due to how genomic information is used. Treatment decisions are often informed by the presence or absence of single mutations in key driver genes; for example, whether or not a patient has a *PIK3CA* mutation or a constitutional pathogenic *BRCA1* or *BRCA2* variant or otherwise.[5,6] Yet, a human cancer genome carries not just one or two causally implicated driver mutations; it carries thousands of mutations, informative of myriad mutational processes that were operative during tumorigenesis, termed mutational signatures.[7–10] Therefore, at present, vast amounts of information present in cancer genomes are not fully utilised in the clinic.

See Online for appendix 1

The whole-genome sequencing (WGS) scalpel-to-report infrastructure offered by the National Health Service (NHS) Genomic Medicine Services evolved out of the national research endeavour (the 100 000 Genomes Project [100kGP]). WGS is already offered on the NHS for paediatric cancers, some haematological conditions, and metastatic diseases. Given these technological and structural advancements, we aimed to assess whether comprehensive WGS linked to mortality data could add prognostic value to existing clinical measures and identify patients with a response to targeted therapeutics to improve breast cancer care.

See Online for appendix 2

Methods

Study design and participants

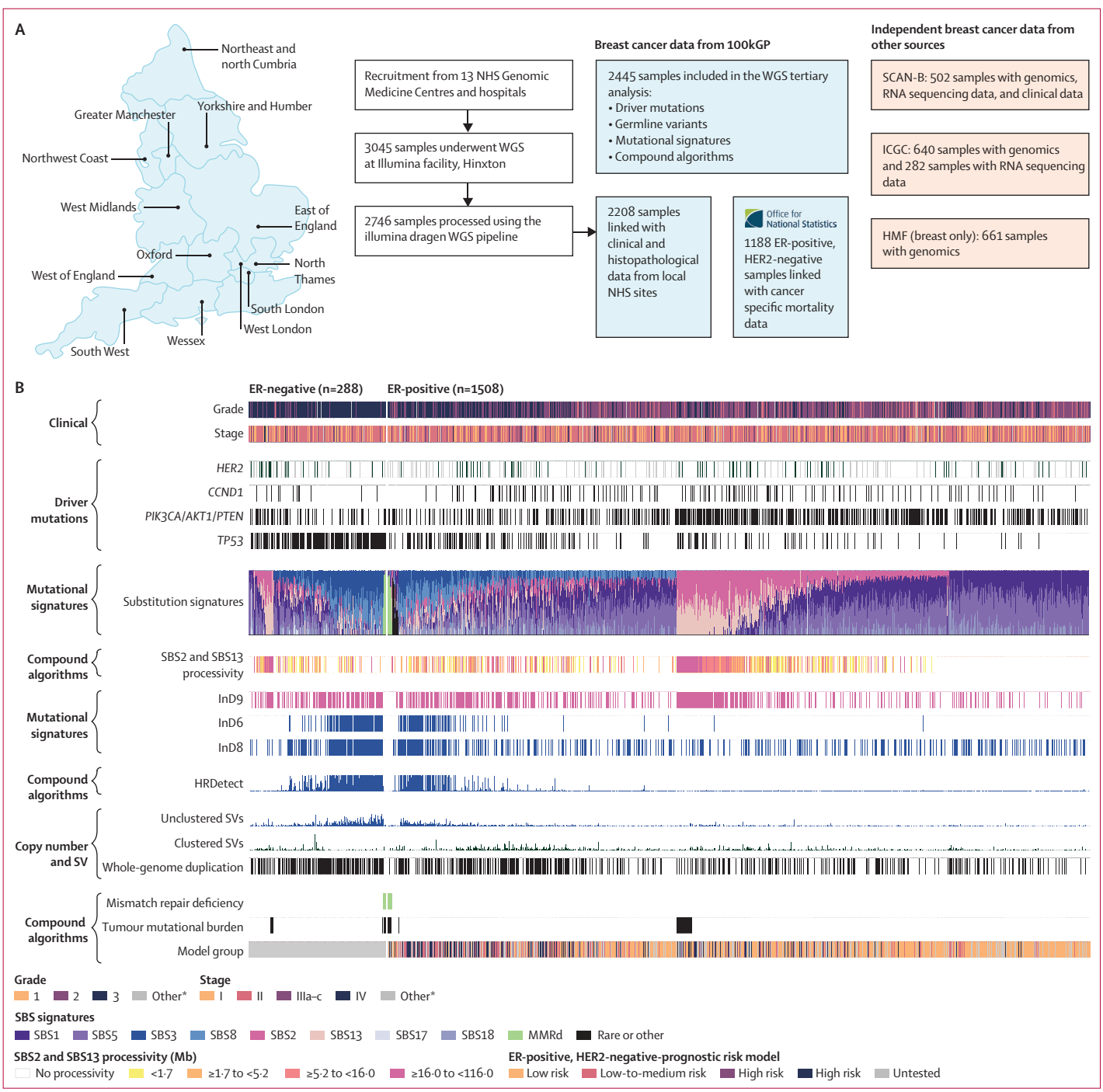
In this integrative, retrospective analysis, we analysed 2445 breast cancer tumours of any stage and molecular subtype collected from 2403 patients. Participants were recruited through 13 NHS Genomic Medicine Centres or hospitals affiliated to the 100kGP between 2012 and 2018 (figure 1A; appendix 1 p 2). There were no specific participant selection criteria. A participant panel advisory group was directly involved in the study design. All participants provided written informed consent.

We linked 2208 (90·3%) tumours from 2204 patients to clinical data, including grade, stage, hormonal receptor status, and age at diagnosis (appendix 2 p 1; table legends for appendix 2 are in appendix 1 pp 31–33). Because of nationally based infrastructure, we could link WGS data to cancer-specific mortality statistics from the UK Office of National Statistics. Data from 1188 patients with stage I–III, ER-positive, HER2-negative breast cancer were linked to mortality data. Records on treatment were too heterogeneous to use in analyses. Sex, race, and ethnicity were not reported in the 100kGP for breast cancer. To validate genomic findings from the 100kGP cohort, we

used data from 1803 additional patients recruited by the Hartwig Medical Foundation (661 patients with metastatic, mixed subtype breast cancer),[11] International Cancer Genome Consortium (640 patients with mixed subtype breast cancer),[12] and Swedish Cancerome Analysis Network-Breast (SCAN-B; 502 patients with ER-positive, HER2-negative breast cancer; appendix 1 p 2).[13] The SCAN-B cohort was used to independently replicate the survival analyses and validate the prognostic framework.

Procedures

Snap-frozen tumour DNA and matched normal DNA from blood samples underwent WGS (mean tumour 96X,



(Figure 1 continues on next page)

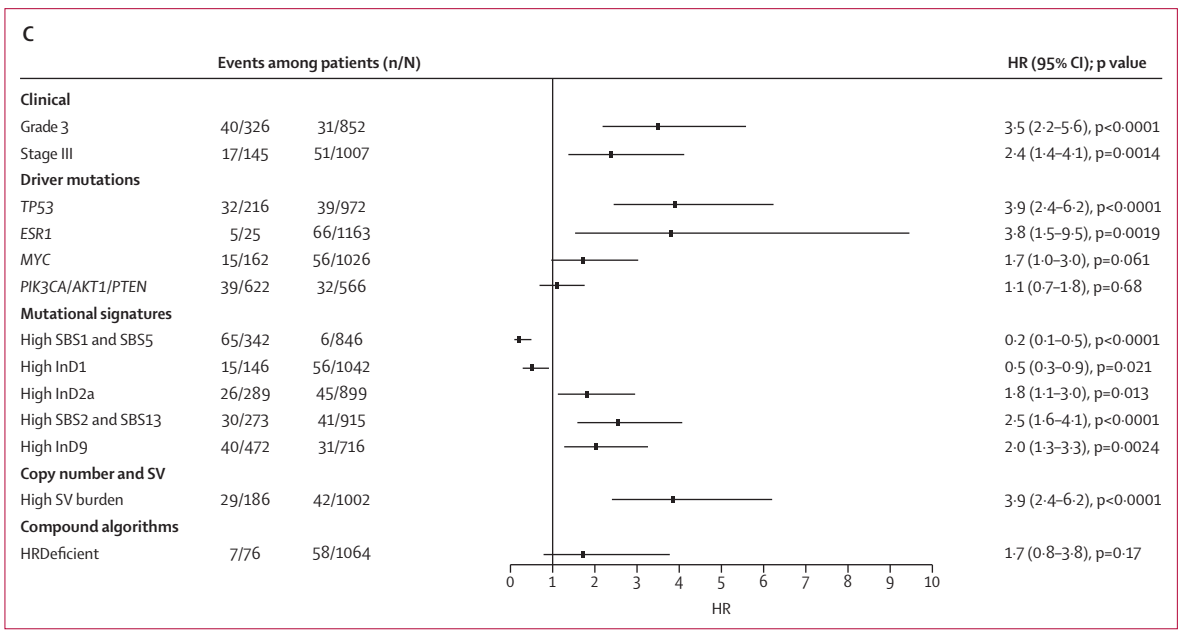


Figure 1: **Overview of selected clinical and genomic characteristics and key prognostic features**
(A) Workflow of national recruitment, sample acquisition, processing, and data generation for the 100kGP cohort. Map reproduced from NHS England, by permission of the Open Government Licence. (B) Clinical and genomic features of ER-negative and ER-positive breast cancer tumours from patients in the 100kGP breast cancer cohort. Processivity refers to the propensity to induce mutations on the same DNA strand over long stretches of DNA. Each tumour is represented on the horizontal axis by stacked vertical bars. Clinical and genomic features are on the left vertical axis. Tumours are ordered by exposure to substitution signatures: rare, associated with homologous recombination repair deficiency, and APOBEC. Where not indicated in the legend, coloured, white, and grey bars indicate presence, absence, or unavailability of data, respectively. (C) Prognostic value of selected genomic features in ER-positive, HER2-negative breast cancer. The number of events and number of patients (left: high or present group; right: low or absent group), HRs (95% CIs), and log-rank p values of univariable Cox regression analysis using the endpoint of cancer-specific mortality are listed. Squares indicate HRs; lines indicate 95% CIs. Detailed survival statistics are presented in appendix 2 p 15. 100kGP=100 000 Genomes project. HMF=Hartwig Medical Foundation. HR=hazard ratio. ICGC=International Cancer Genome Consortium. InD=insertion and deletion. NHS=National Health Service. SBS=single-base substitution. SCAN-B=Sweden Cancerome Analysis Network-Breast. SV=structural variation. WGS=whole-genome sequencing. *Including no data.

matched normal 36X; appendix 1 pp 2–3). The Genomics England 100kGP core bioinformatics pipelines were used to obtain high-quality data.[14] Between September, 2022, and January, 2025, we performed comprehensive WGS analyses seeking somatic or constitutional driver mutations, mutational signatures, and compound algorithmic scores for homologous recombination repair deficiency (HRD), mismatch repair deficiency, and tumour mutational burden.[10,15,16]

To identify tumours with HRD, a biomarker for selective sensitivity to DNA damaging agents and PARP inhibitors, we assessed the genomes of tumours for characteristic mutational scars that have previously been linked to HRD.[17] Tumours with *BRCA1* or *BRCA2* deficiency have distinct HRD mutational signatures: single-base substitution (SBS) patterns (SBS3 and SBS8), structural variation (SV) signatures defined by dispersed tandem duplications of less than 10 kb (R3, associated with *BRCA1* loss), deletions of less than 10 kb (R5, associated with *BRCA2* loss), small deletions with microhomology, and extensive copy number losses.[12] These genomic signatures have been condensed into a machine learning score called HRDetect, which we applied to all tumours.[15]

Similarly, to identify mismatch repair (MMR) deficiency, an indicator of response to immunotherapies in many cancer types,[18] we used the PRRDetect R package to identify tumours with characteristic substitution and indel signatures of MMR deficiency or polymerase dysfunction.[10] Tumour mutational burden, a proxy for mismatch repair deficiency, was also calculated for all samples.[19]

To assess whether genomic features are valuable indicators of outcome, we systematically surveyed WGS features seeking prognosticators that could complement existing clinical measures, highlighting mechanistic explanations where possible. We focused on mortality associations in patients with ER-positive, HER2-negative breast cancer in the 100kGP cohort, which was most adequately powered for this analysis (n=1188). Median patient follow-up was 4·6 years (95% CI 4·6–4·7). We used Cox regression multivariable analysis to identify genomic features that provided prognostic value independently of each other and combined these into a prognostic model intended to be used alongside existing clinical measures.

To validate the prognostic model, we applied it to 502 ER-positive, HER2-negative breast cancers in the SCAN-B cohort. WGS tumours from the SCAN-B cohort were sequenced to a lower depth of 36X (*vs* 96X in the primary cohort). To ensure generalisability of our

framework across various sequencing depths, we performed in-silico down sampling and complete WGS reanalysis on 46 tumours from the 100kGP breast cancer cohort. Relevant to the factor of SV burden, we produced an estimate of equivalent SV burden thresholds across the range of depths. Furthermore, using the SCAN-B cohort, we compared our prognosticator to multigene expression predictors favoured in high-income countries and institutions, including Oncotype DX (Exact Sciences; Madison, WI, USA) and MammaPrint (Agendia; Irvine, CA, USA).[20] Detailed methods are provided in the appendix (pp 3–12).

Outcomes

The primary outcomes of this study were to identify strongly prognostic WGS features that are independently informative of existing clinical markers in patients with ER-positive breast cancer, as well as predictive WGS features that can be used to triage patients for targeted therapy or clinical trial recruitment. Secondary outcomes were the development of a WGS prognostic framework for clinical application, intended to complement existing prognostic measures, and the identification of novel breast cancer biology through mechanistic analysis of mutational patterns.

Statistical analysis

To evaluate the prognostic value of WGS features, we performed univariable and multivariable Cox regression with a cancer-specific mortality endpoint on data from patients with stage I–III, ER-positive, HER2-negative breast cancer in the 100kGP cohort (R survival [version 3.8.3]). Mortality not related to breast cancer was treated as a censored observation. Results were considered significant at a p value of less than 0·05 and the proportional hazard assumption was assessed visually. For omics analyses, driver–event enrichment were tested with Fisher's exact tests, expression differences between subgroups with Wilcoxon *t* tests, and signature correlations with DNAse I hypersensitivity sites with Pearson's correlation coefficients. Modelling of SV burden by sequencing depth was conducted with linear mixed-effects models (lme4, core R package [version 1.1.37]).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

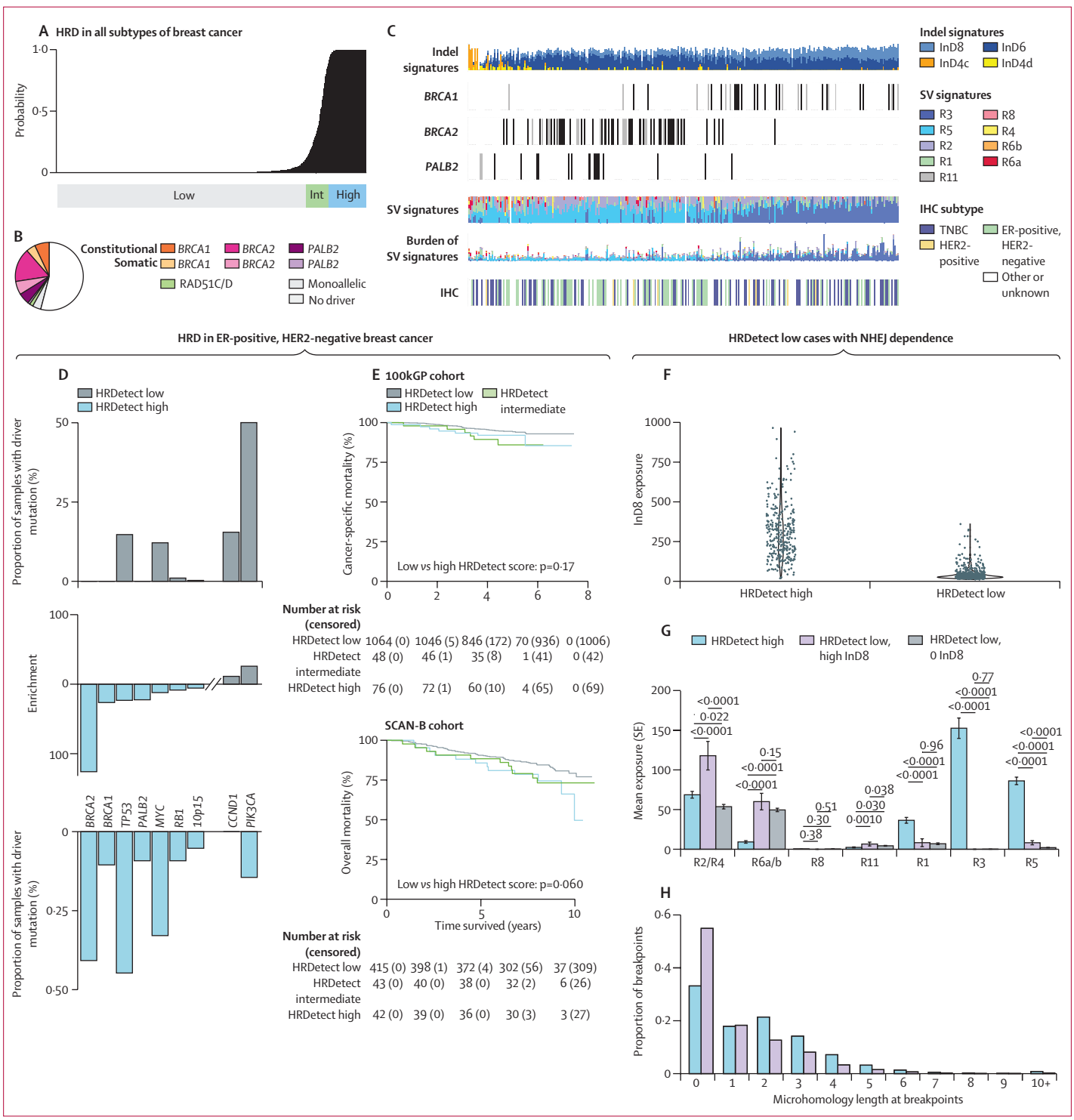
Results

The 100kGP breast cancer cohort showed characteristics for cancer stage, tumour subtype, and age at diagnosis, representative of the UK population; however, we noted a depletion of low-grade tumours (8·7% observed *vs* 17·0% expected; appendix 1 pp 13–14).[21] The landscape of driver mutations, mutational signatures, and copy number profiles were consistent with previous reports (appendix 1 pp 13–19; appendix 2 pp 2–9).[12,22] We saw typical copy number gains of 1q, 8q, and 17q, and loss of heterozygosity of 1p and 16q.[23] Whole-genome duplication occurred in 1072 (43·8%) of 2445 tumours with early or mid-cancer evolution (median mutational time 0·44 [IQR 0·22–0·65]; appendix 2 p 10). *PIK3CA* driver mutations were most frequent (995 [40·7%]), followed by *TP53* driver mutations (731 [29·9%]; appendix 1 pp 13–14). Other frequent drivers included amplifications of *MYC*/8q24, *CCND1*/11q13.3, *FGFR1*/8p11, *ERBB2*/17q23, 1q32, *PAK1*/11q13–14, 17q23, and *GNAS*/20q13, as well as *CDH1, GATA3*, and *PTEN* mutations (appendix 1 pp 13–14; appendix 2 p 2). 253 (6·9%) of 3643 small variant drivers were sub-clonal, mainly in *PIK3CA* and *TP53* (appendix 2 p 11). A single *MYB–NFIB* gene fusion in an adenoid cystic carcinoma and individual *VIT1A–TCF10* and *ETV6–NTRK3* fusions were identified.[24] SVs involving *ESR1* were noted (appendix 2 p 3). We identified a small number of novel breast cancer genes, but no novel, recurrent, non-coding driver mutations. Hypermutable non-coding sites were noted as observed previously (appendix 1 p 12–13).[12]

298 (12·2%) of all 2445 tumours showed a high HRDetect score (>0·9), indicating HRD; 1959 (80·1%) had a low score (<0·1); and 188 (7·7%) had an intermediate score (0·1–0·9; figure 2A). Of the 298 tumours with a high HRDetect score, 120 (40·2%) had a somatic or constitutional driver in a canonical *HR* gene, with concomitant loss of the wild-type parental allele (figure 2B, C). Methylation data were not available for investigating the underlying HRD inactivation mechanism in the remaining 178 (59·7%) tumours. Although HRD is usually associated with triple-negative breast cancer, 76 (25·5%) tumours with high HRDetect scores were ER-positive, HER2-negative. Eight (4·7%) of 171 HER2-positive cases also showed HRD. ER-positive, HER2-negative tumours with a high HRDetect score showed triple-negative breast cancer-like features (eg, enrichment of *TP53, RB1,* and *MYC*; figure 2D; appendix 1 p 20; appendix 2 p 14). 23 (30·3%) of the 76 ER-positive, HER2-negative tumours with a high HRDetect score had a *BRCA1*-like phenotype.

Despite the trend to improved time to relapse and improved survival in HRD cases among patients with triple-negative breast cancer,[17] we observed the opposite trend in HRD cases among patients with ER-positive, HER2-negative breast cancer. However, this trend did not reach significance due to low numbers of HRD cases in the ER-positive, HER2-negative breast cancer group (hazard ratio [HR] 1·7 [95% CI 0·8–3·8]; p=0·17; figure 2E; appendix 2 pp 15–16). The trend towards poor survival in patients with HRD in the ER-positive, HER2-negative breast cancer subtype was replicated in the independent SCAN-B cohort (1·8 [1·0–3·3]; p=0·060; figure 2E) and is in keeping with findings in ER-positive

BRCA1 or *BRCA2* mutation carriers.[25] This finding was reinforced by comparing the frequency of HRD in 661 patients with metastatic breast cancer from the Hartwig Medical Foundation with cases of invasive breast cancers in the 100kGP cohort. Although triple-negative breast cancers with a high HRDetect score were depleted in metastatic breast cancer (26 [32·1%] of 81) compared with invasive breast cancer (86 [50·3%] of 171), ER-positive, HER2-negative cases with a high HRDetect score were higher among individuals with metastatic breast cancer (46 [10·8%] of 424) than among those with invasive breast cancer (76 [6·4%] of 1191), suggesting an



enrichment of potentially under-diagnosed, inadequately treated patients with HRD in the ER-positive, HER2-negative breast cancer subtype progressing to metastatic disease.

Mutational signatures could further distinguish subtypes within HRD, differentiating *BRCA1*-like and *BRCA2*-like cancers by SV signatures (R3 *vs* R5) and indel signature; InD6 (deletions at microhomology); and InD8, characterised by indels with at least 5 bp with little to no microhomology, attributed to non-homologous end joining (figure 2C). Indel signatures could further distinguish a 2% subset of HRD cancers by an alternative signature, InD4c, linked to *TOP1*-related mutagenesis associated with transcription.[26] Cancers with a high HRDetect score and indel signature InD4c did not have canonical HR drivers, showed longer tandem duplication SVs, and had rare occurrences of biallelic loss of the helicase *SETX* (appendix 1 p 20; appendix 2 p 17–18). Further work is required to understand the cause and therapeutic vulnerabilities of this patient subgroup who are distinguishable by these biomarkers.

Other biological abnormalities that increase single-strand breaks have been postulated to be sensitive to PARP inhibition due to increased PARP dependency.[27] We identified 87 (3·6%) of 2445 tumours with glycosylase abnormalities that could be considered compromised in base excision repair. One patient in this group had SBS30 caused by biallelic *NTHL1* loss and the remaining 86 tumours had outlier SBS18 exposure (associated with compromised *OGG1* activity). Additionally, cancers with an intermediate HRDetect score are a diverse group that might have a similar susceptibility to DNA damaging agents or PARP inhibition. Tumours with an intermediate score showed an enrichment for SV signature R1 (long tandem duplications >100 kb to 1 Mb) and *CCNE1* amplification (appendix 2 p 19). Additionally, compared with tumours with a low HRDetect score, those with an intermediate score showed a trend towards worse outcomes in patients with ER-positive, HER2-negative breast cancer (HR 2·4 [95% CI 1·0–5·6]; p=0·035; figure 2E). Finally, we observed cases of breast cancer with an unexplained dependence on non-homologous end joining (figure 2F–H), highlighting another distinguishable patient subpopulation for potential clinical studies.

Mismatch repair deficiency is an indicator of response to immunotherapies in many cancer types, but is not routinely tested for in breast cancer (appendix 2 p 20).[18] Tumour mutational burden is used as a proxy for mismatch repair deficiency, with mixed results in breast cancer.[19] Among all 2445 breast cancers, 93 (3·8%) had high tumour mutational burden, with diverse causes including APOBEC activity (63 [67·7%]) and mismatch repair deficiency (13 [14·0%]; appendix 1 p 21). Therefore, tumour mutational burden was not a specific biomarker (80 [86·0%] false positives) nor was it prognostic (HR 1·1 [95% CI 0·27–4·4]; p=0·91; appendix 2 p 15). Using the PRR Detect algorithm,[10] we found mismatch repair deficiency in 16 (0·7%) of 2445 tumours (appendix 2 p 10). Nine (56·3%) of these tumours had biallelic loss of an *MMR* gene (ie, *PMS2*, *MSH6*, *MSH2*, or *MLH1*). Of 15 breast cancers with mismatch repair deficiency and linked clinical annotation, nine (60·0%) were ER-positive and 14 (93·3%) were stage I or II. Therefore, these tumours would be missed under current eligibility for immunotherapy in patients with triple-negative breast cancer. Other genomic abnormalities not tested for routinely in breast cancer, which are actionable in other organs, include driver mutations in *KRAS* (12 [0·5%] of 2445 tumours), *EGFR* (50 [2·0%]), *CCNE1* (52 [2·1%]), and *BRAF* (four [0·2%]; appendix 2 p 20). These individualities highlight patient subsets warranting prospective clinical studies to gather evidence supporting genome-directed tumour-agnostic therapy.

Beyond targetable features, WGS can identify resistance markers. *ESR1* driver mutations are implicated in resistance to endocrine therapy.[28] 33 (2·2%) of 1490 cases of ER-positive invasive breast cancer in the 100kGP cohort showed *ESR1* drivers: five gene fusions, 12 small variants (two subclonal), and 16 amplifications. These drivers were associated with poor outcomes in patients with ER-positive, HER2-negative breast cancer (HR 3·8 [95% CI 1·5–9·5]; p=0·0019; appendix 1 p 22). A further 28 (1·9%) ER-positive invasive breast cancers had SVs involving *ESR1* (appendix 2 p 3).[29] It is unclear whether all such events conferred resistance; however, patients in this group also had worse outcomes than those with other breast cancers (HR 2·8 [95% CI 0·89–9·0]; p=0·066).[29] If all *ESR1* drivers and SVs are potentially informative,

Figure 2: **Recombination and repair defects in breast cancer**
(A) HRDetect scores of 2445 breast cancer tumours from the 100kGP breast cancer cohort. The following HRDetect category boundaries are indicated below the graph: low (<0·1), intermediate (≥0·1 to <0·9), and high (≥0·9). (B) Identification of causative driver mutation underlying the HRD phenotype in 298 cases of breast cancer with high HRDetect scores. (C) Mutational signature exposure, biallelic driver mutations (somatic [grey] and constitutional [black]), and immunohistochemical subtype of tumours with high HRDetect scores. Tumours are ordered by the proportion of indel mutations assigned to InD8. (D) Driver mutation enrichment in ER-positive, HER2-negative tumours with high versus low HRDetect scores were calculated with Fisher's exact tests (log2[q]). The central panel shows the direction of enrichment. Only significantly enriched genes are shown. (E) Kaplan–Meier analysis of patients with ER-positive, HER2-negative breast cancer stratified by HRDetect group (low, intermediate, and high); high versus low HRDetect scores in the 100kGP breast cancer cohort (HR 1·7 [95% CI 0·8–3·8]; p=0·17) and the SCAN-B cohort (1·8 [1·0–3·3]; p=0·060). (F, G, H) 58 (3·0%) of 1959 tumours with low HRDetect scores had an unexplained dependence on NHEJ, as characterised by a high exposure to NHEJ-associated indel signature, InD8 (F), and the occurrence of translocations (R2/R4) in the subgroup with low HRDetect scores and high InD8 exposure (p values from Wilcoxon tests; G) and limited to no microhomology at translocation breakpoint junctures (H; KS-test D=0·52; p=0·0004), indicating blunt end repair associated with NHEJ rather than a microhomology-mediated repair pathway. The plot shows microhomology length at characteristic SV breakpoints of tumours with high HRDetect scores (non-clustered tandem duplications and deletions) compared with tumours with low HRDetect scores and high levels of InD8 (translocations). HRD=homologous recombination repair deficiency. IHC=immunohistochemistry. NHEJ=non-homologous end joining. SV=structural variation. TNBC=triple-negative breast cancer.

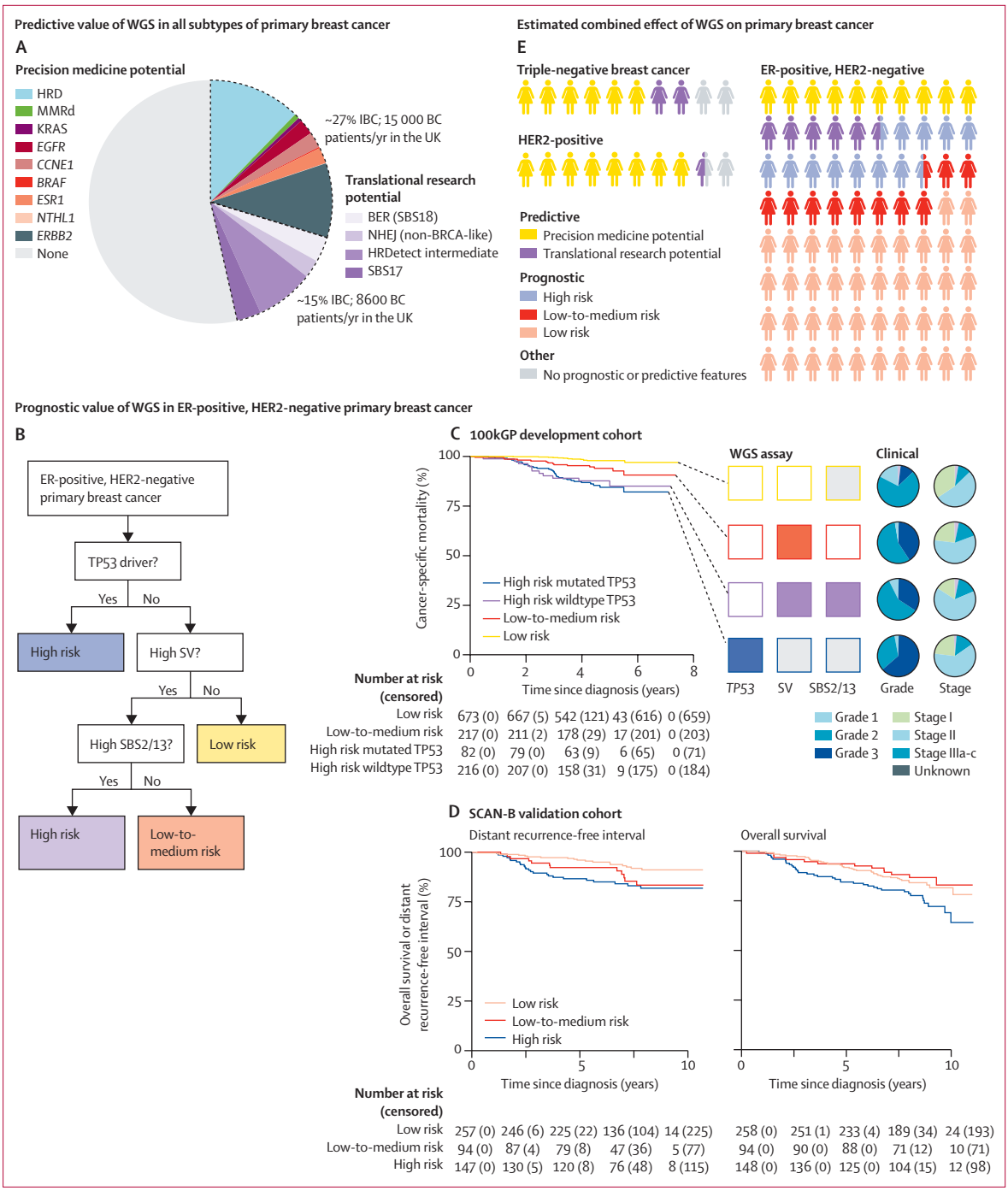


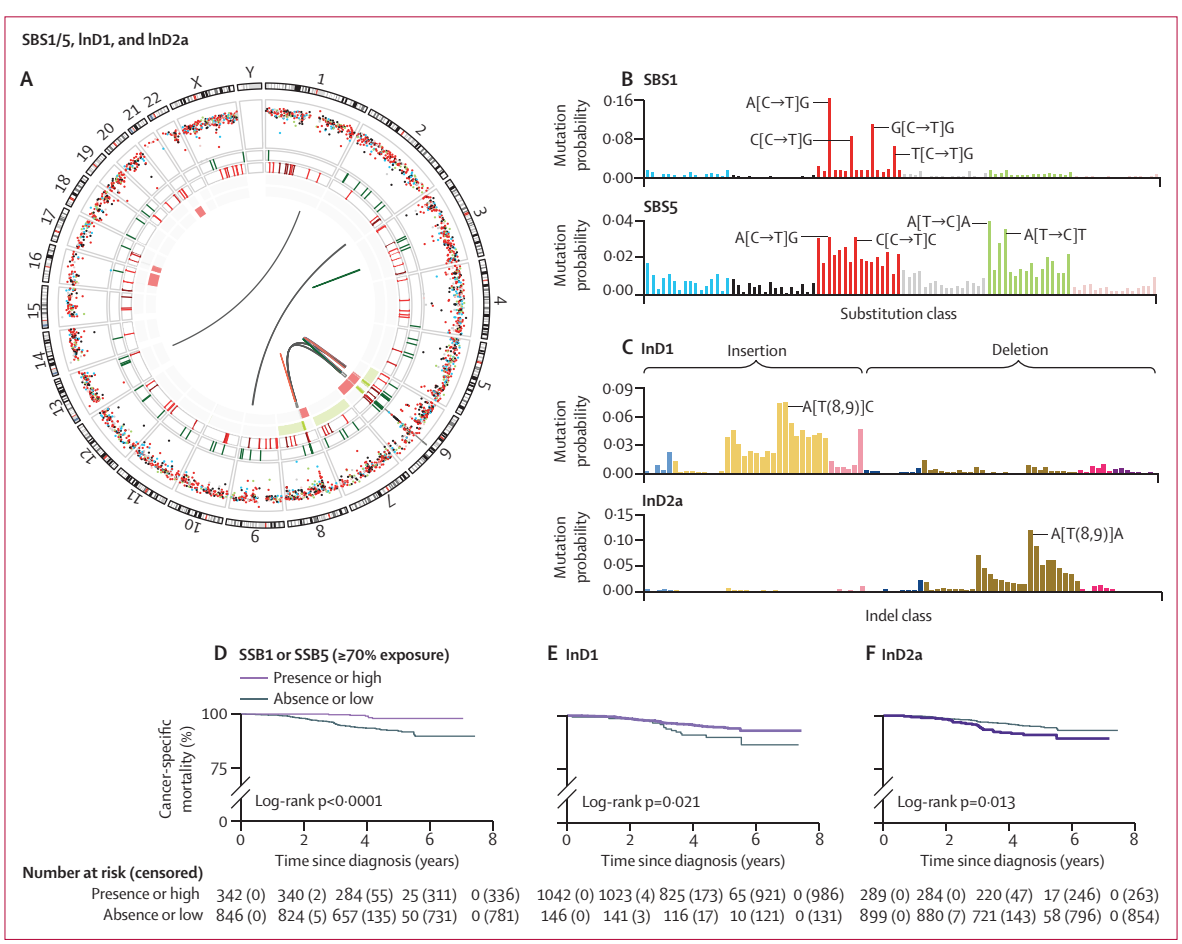
Figure 3: **WGS-based stratification of breast cancer**

(A) Predictive utility of WGS for response to targeted therapeutics in all breast cancer subtypes, excluding stage IV disease. WGS features with precision medicine potential (yellow group) or translational research potential (purple group) are indicated. Grey indicates a tumour in which neither a precision medicine potential nor a translational research potential feature was identified. Estimated impact of triaging and treating patient by features with precision medicine or translational research potential in the UK based on 36 800 new diagnoses per year.[30] (B) Overview of prognostic framework for ER-positive, HER2-negative breast cancer. (C) Cancer-specific survival for each genomic risk group in the 100kGP development cohort of patients with ER-positive, HER2-negative breast cancer. Dotted lines link each curve to the genomic features (*TP53* driver mutation, SV count >90, and SBS2 or SBS13 exposure >25%) and clinical features (grade and stage) for each risk group. Filled, open, and grey squares indicate whether the group is positive, negative, or mixed for the genomic feature indicated, respectively. Pie charts show tumour grade and stage distribution per risk group. (D) Application of the prognostic framework to SCAN-B validation cohort. Endpoints are distant recurrence-free interval and overall survival at 10 years. (E) Schematic representation of 100 patients with invasive breast cancer grouped by immunohistochemical subtype indicating the proportion with at least one predictive feature and, for ER-positive, HER2-negative breast cancer, prognostic WGS features. BC=breast cancer. HRD= homologous recombination repair deficiency. HRDNHEJ=non-homologous end joining. IBC=invasive breast cancer. MMRd=mismatch repair deficient. SV=structural variation. WGS=whole-genome sequencing.

62 (4·1%) of all 1508 women with ER-positive breast cancer (any stage) in the 100kGP cohort would be affected.

To summarise known and new precision medicine potential of WGS in breast cancer, 645 (26·7%) of 2412 invasive breast cancers (excluding stage IV cancers) in the 100kGP cohort had a genomic feature with either immediate actionability (eg, *ERBB2* 238 [9·8%]), clinical trial potential (HRD 295 [12·2%]), mismatch repair deficiency (16 [0·7%]), a driver actionable in other organs (*NTHL1, CCNE1, BRAF, KRAS*, or *EGFR*; 110 [4·6%]), or a marker of resistance (*ESR1* drivers, ER-positive only 33 [2·2%] of 1490; figure 3A; appendix 2 p 20). We estimate that 364 (15%) of 2412 invasive breast cancers harbour mutational signatures offering potential for translational research (figure 3A–E).

As well as offering predictive utility, a systematic survey of WGS features yielded several prognosticators that could add value to the existing clinical measures used to stratify patients with ER-positive, HER2-negative breast cancer by risk in the UK (appendix 1 pp 23–24). The most prevalent SBS signatures were benign mutational processes—namely, SBS1, the signature of spontaneous deamination of 5-methylcytosine (1985 [81·2%] of 2445), and the commonly co-assigned SBS5 (cause unknown; 1840 [75·3%]; appendix 1 p 13). A high proportion of mutations assigned to SBS1 and SBS5 was associated with clinically and genomically favourable features (figure 4A, B; appendix 1 p 25). The indel signature, InD1, characterised by 1 bp T insertion at polynucleotide tracts (>5 nt) and postulated to be caused by replication-related nascent strand slippage, was also ubiquitous (prevalence 2046 [83·7%] of 2445 tumours; figure 4C; appendix 1 p 25). Unexpectedly, the signature thought to represent the equivalent process on the template strand, InD2a, was present in fewer tumours (prevalence 732 [29·9%]; figure 4C). Investigating further, the differential prevalence of InD2a and InD1 was also observed in other tumour types and is, therefore, not restricted to breast cancer. InD2a prevalence was higher in tumour types and breast cancer subtypes with higher proliferation (appendix 1 p 25). Additionally, among 500 ER-positive, HER2-negative breast cancer tumours from the SCAN-B cohort and 280 from the International Cancer Genome Consortium with linked transcriptomic data, cases assigned to InD2a had elevated transcriptional proliferation markers (*MKI67, MCM2*, and *PCNA*;



(Figure 4 continues on next page)

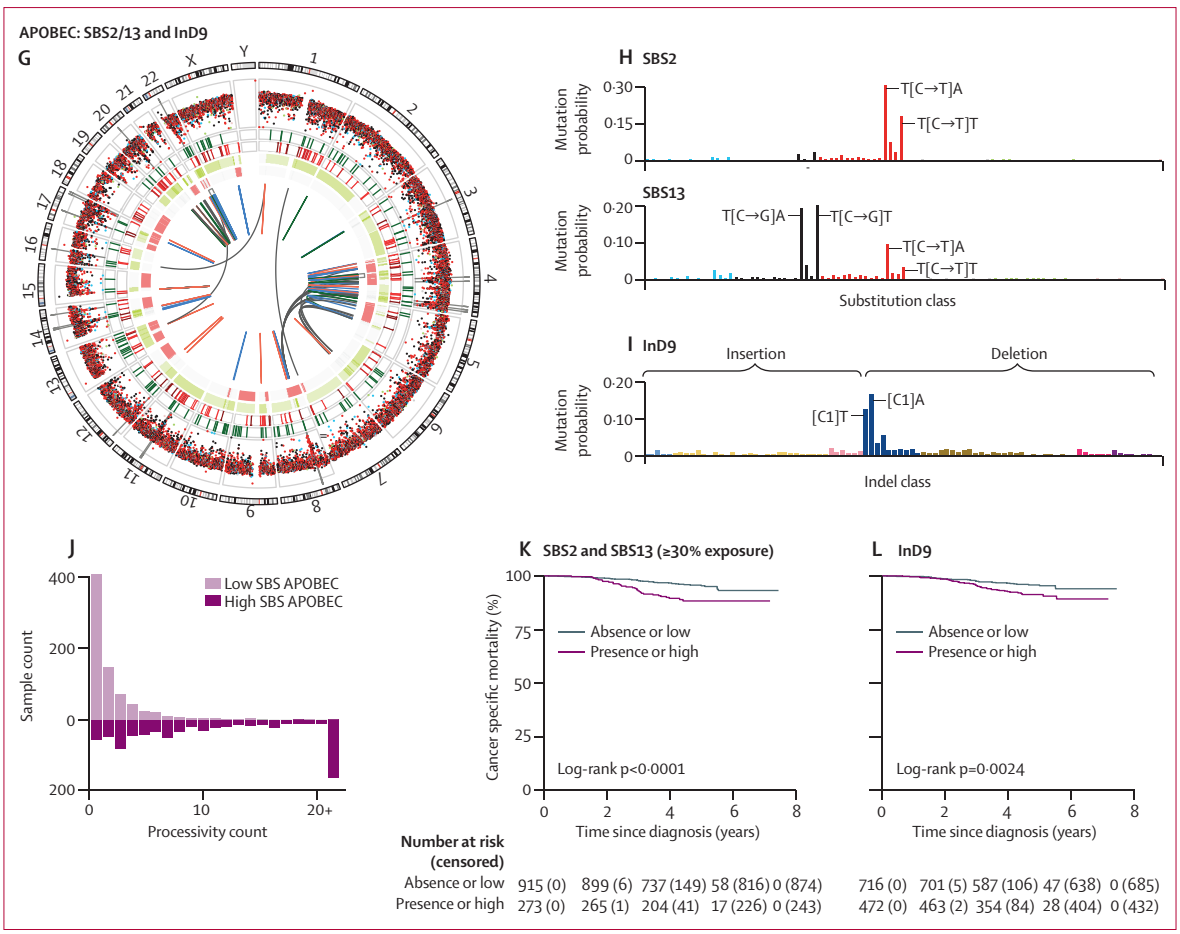


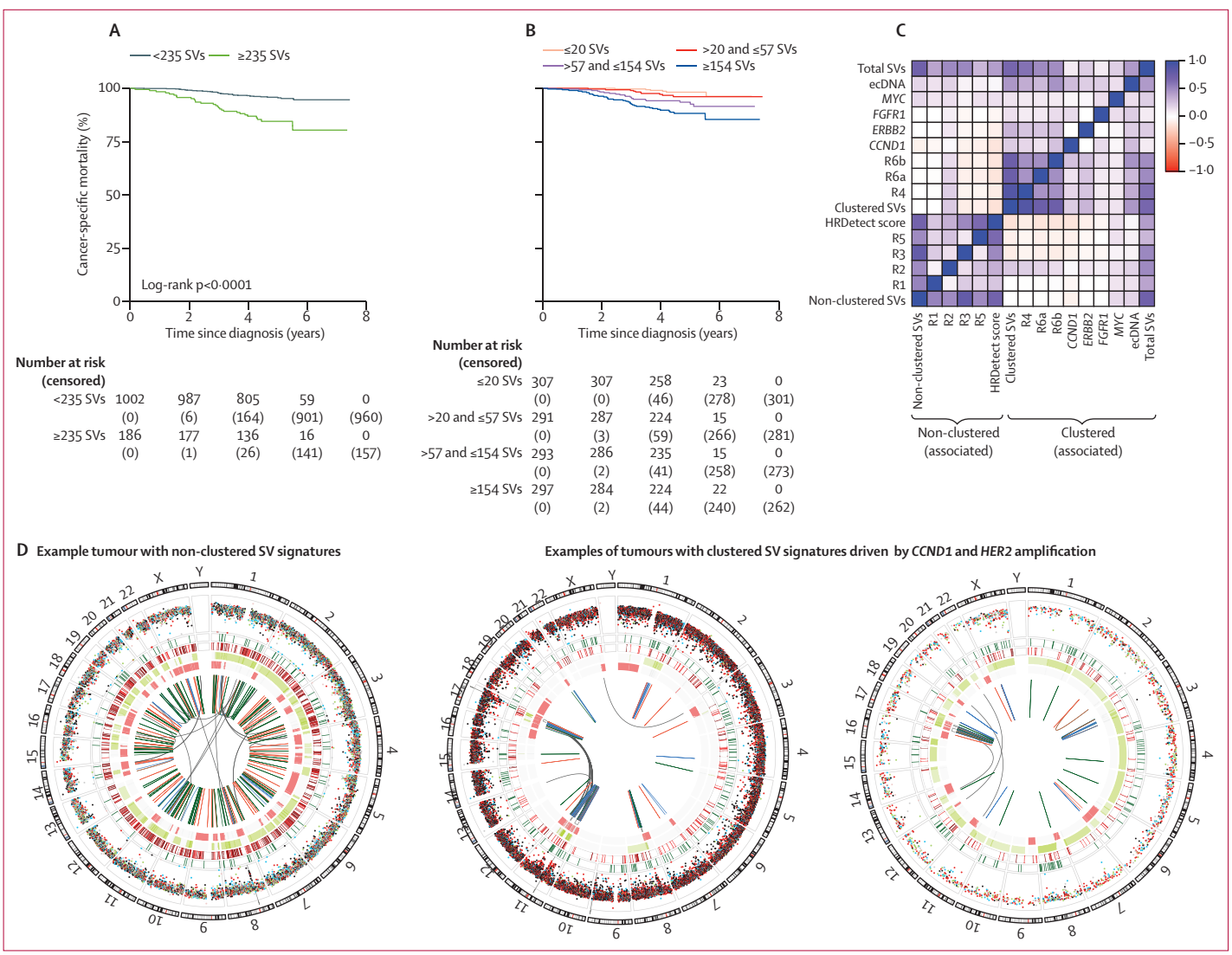
Figure 4: **Mechanistic insights into endogenous mutational processes and associations with cancer-specific mortality**
(A) Whole-genome profile of a typical tumour with mainly benign endogenous mutational processes. Genomic features depicted in a circos plot from outermost rings inwards: karyotypic ideogram; substitutions plotted as log10 inter-mutation distance on radial axis (C→A [blue], C→G [black], C→T [red], T→A [grey], T→C [green], and T→G [pink]); small insertions and deletions (ring with short green and red lines); total (gain [>2; green]) and minor (loss [red]) copy number; and SVs (central lines: tandem duplications [green]; deletions [red]; inversions [blue]; and translocations [grey]). (B) Mutational signatures SBS1 and SBS5. (C) Mutational signatures InD1 and InD2a. A description of the indel catalogue is provided in appendix 1 (pp 18–19). (D, E, F) Kaplan–Meier analyses of cancer-specific mortality comparing high versus low SBS1 and SBS5 exposure (≥70%; D) and the presence versus absence of InD1 (E) and InD2a (F) in patients with ER-positive, HER2-negative breast cancer. (G) Whole-genome profile of an APOBEC-mutagenised tumour. (H) APOBEC-related substitution signatures. (I) APOBEC-related indel signature. (J) Processivity count in tumours with high (median or above) versus low APOBEC exposures. (K, L) Kaplan–Meier analyses of cancer-specific survival comparing high versus low SBS2 and SBS13 exposure (≥30%; K) and the presence versus absence of InD9 (L). InD=insertion and deletion. SBS=single-base substitution. SV=structural variation.

appendix 1 p 25) and Ki-67 staining (odds ratio 5·8; p=0·0006; n=336 in the SCAN-B cohort). The clinical significance of discerning between these signatures is underscored by the finding that tumours with a high proportion of SBS1 and SBS5 (HR 0·22 [95% CI 0·09–0·50; p<0·0001; figure 4D) and tumours assigned InD1 (0·52 [0·29–0·91]; p=0·021; figure 4E) were associated with improved outcomes, independent of customary clinical prognosticators. By contrast, InD2a was enriched in high-grade cancers and modestly associated with poor survival in patients with ER-positive, HER2-negative breast cancer, in keeping with lnD2a being an indicator of uncontrolled proliferation (1·80 [1·10–3·00]; p=0·013; figure 4F).

Substitution signatures related to APOBEC are characterised by C→T transitions (SBS2) and C→G transversions (SBS13) preceded by a 5'T (figure 4G, H). APOBECs are associated with a 1 bp C deletion signature (InD9), also preceded by a 5'T (figure 4I).[10] Most tumours had a small contribution of APOBEC-related signatures (1873 [76·6%] of 2445). A minority defined by an outlier analysis (349 [14·3%]) showed extensive substitution mutagenesis and were termed quantitatively and qualitatively a hypermutator phenotype, in which SBS APOBEC (SBS2 and SBS13) and InD9 were strongly correlated (r=0·72; p<0·0001). The hypermutator SBS2 and SBS13 phenotype was associated with a processive quality: the propensity to induce mutations on the same DNA strand over long stretches of DNA (figure 4J). Processive events spanned impressive genomic lengths and occurred more frequently on chromosome arms with an increased

frequency of DNase I hypersensitive sites (figure 1B, appendix 2 p 21). Therefore, chromatin openness might influence the likelihood of high-level substitution mutagenesis caused by APOBECs. Processivity was not, however, observed for APOBEC indels. This finding, combined with differences in relative prevalence by subtype of breast cancer, differential correlations with SV types, and the presence of breast cancer cases with high levels of InD9 without SBS2 or SBS13 point to different mechanistic triggers (appendix 1 p 26). From a clinical perspective, APOBEC signatures were associated with poor outcomes independently of customary clinical features (HR 2·5 [95% CI 1·6–4·1]; p<0·0001 for high-level SBS2 or SBS13 and 2·0 [1·3–3·3]; p=0·0024 for InD9; figure 4K, L). The correlated observation of localised hypermutation, termed kataegis, was also associated with poor survival (HR 3·2 [95% CI 1·8–5·7; p<0·0001; appendix 1 p 24).

Beyond substitution and indel patterns, total SV burden was a strong independent predictor of poor outcomes in patients with ER-positive, HER2-negative breast cancer (HR 3·9 [95% CI 2·4–6·2]; p<0·0001; figure 5A; appendix 1 p 27). SV burden was a stronger prognosticator than grade (figure 1C), reinforced by a progressive relationship between burden and outcomes (figure 5B). SV burden encompassed a broad spectrum of structural mutagenic processes, from focal oncogene amplifications (eg, *ERBB2* or *CCND1*) to dispersed SVs in HRD (figure 5C, D). We investigated SV signatures (appendix 1 p 17; appendix 2 p 5) and other SV-associated features (eg, amplification of oncogenes and presence of extrachromosomal DNA[31]) individually as prognosticators. We found that, although each feature was negatively prognostic, none prognosticated as well as total SV burden (figure 5E). This finding is likely because SV features are collinear with each other; therefore, these



(Figure 5 continues on next page)

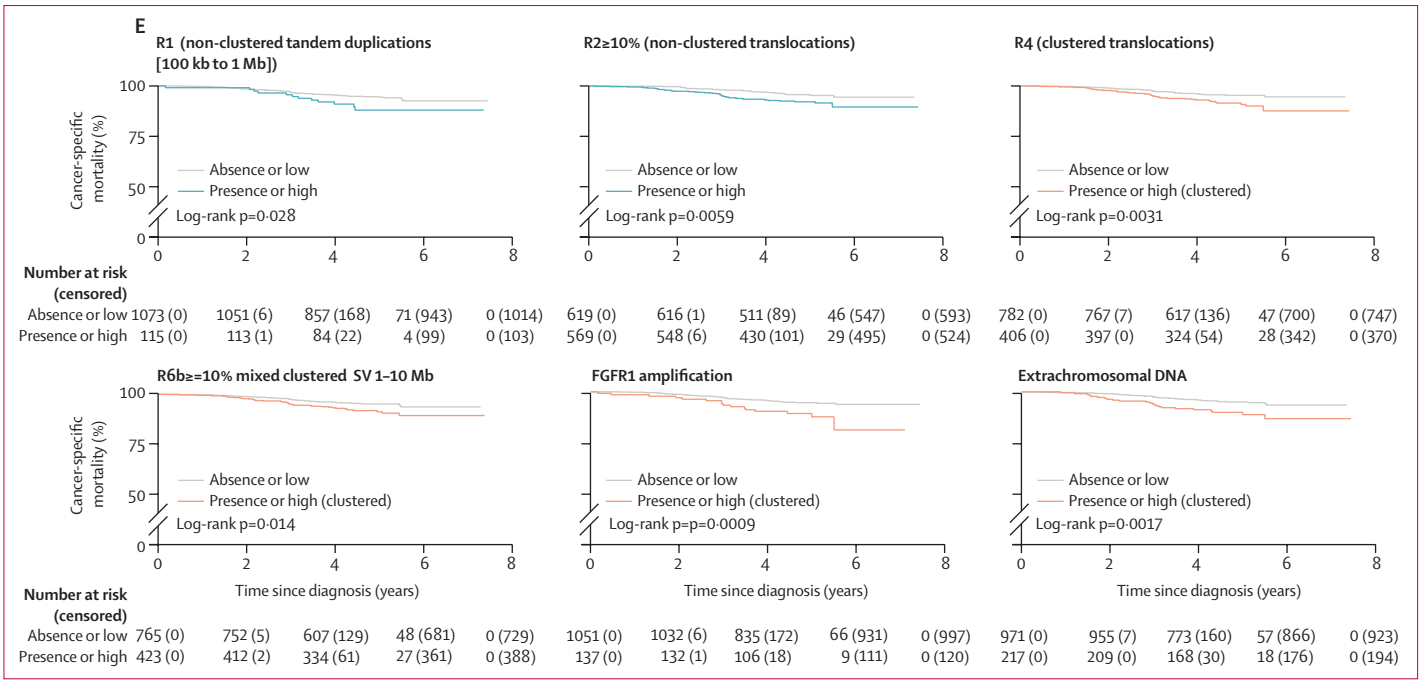


Figure 5: **Associations between SV and cancer-specific mortality**

(A) Kaplan–Meier analysis of cancer-specific mortality comparing high (≥235 SVs) versus low (<235 SVs) SV burden in 1188 patients with ER-positive HER2-negative breast cancer. (B) Kaplan–Meier analysis of cancer-specific mortality comparing low to high SV quartiles. (C) Correlations between SV-associated features. The scale represents Pearson's R coefficient. (D) Whole-genome profiling of three tumours with high SV burden capturing different underlying biology: HRD, *CCND1*-amplified, and *HER2*-amplified tumours (left to right). Genomic features depicted in a circos plot from outermost rings inwards: karyotypic ideogram; substitutions plotted as log10 inter-mutation distance on radial axis (C→A [blue], C→G [black], C→T [red], T→A [grey], T→C [green], and T→G [pink]); small insertions and deletions (ring with short green and red lines); total (gain [>2; green]) and minor (loss [red]) copy number; and SVs (central lines: tandem duplications [green]; deletions [red]; inversions [blue]; and translocations [grey]). (E) Kaplan–Meier analysis of selected genomic features contributing to high SV burden (R1, R2, R4, R6b, FGFR1 amplification, and extrachromosomal DNA). Grey lines indicate tumours with a low burden of SV signatures or in the case of FGFR1 amplification or extrachromosomal DNA, which are negative for these features. Coloured lines indicate tumours that have a high burden of SV signatures or are positive for FGFR1 amplification or extrachromosomal DNA (blue for non-clustered SV-associated features and orange for clustered SV-associated features). All data are provided in appendix 2 p 15. ecDNA=extrachromosomal DNA. SV=structural variation.

features were affected by reduced statistical power when tested individually (figure 5C). Given that double strand breaks are not well tolerated by normal cells, extensive SVs might serve as an insignia of markedly abnormal cells. Our findings are corroborated by a 2025 study, which noted high SV in transcriptome-defined groups with poor survival.[32]

Among driver events in ER-positive, HER2-negative breast cancer, *TP53* mutations were an independent prognosticator superior to histological grade (HR 3·9 [95% CI 2·4–6·2]; p<0·0001; appendix 1 p 22). Although *RB1* mutations, typically associated with triple-negative breast cancer, were infrequent in patients with ER-positive, HER2-negative breast cancer (20 [1·7%] of 1203 tumours), they were associated with poor prognosis (HR 5·0 [95% CI 2·0–13·0]; p=0·0001). Mutations in *PIK3CA*, *AKT1*, and *PTEN* were not prognostic (1·1 [0·7–1·8]; p=0·68), in line with previous reports.[33]

To summarise, the strongest prognosticators of adverse outcomes in ER-positive, HER2-negative invasive breast cancer were burden of SVs (HR 3·9 [95% CI 2·4–6·2]; p<0·0001), *TP53* drivers (3·9 [2·4–6·2]; p<0·0001), and high levels of APOBEC-related mutagenesis (SBS2 and SBS13; 2·5 [1·6–4·1]; p<0·0001). Dominance of SBS1 and SBS5 (0·22 [0·09–0·50]; p<0·0001) and indel signature InD1 (0·52 [0·29–0·91]; p=0·021) were associated with improved outcomes (figure 1C). All features remained prognostic after adjusting for grade, stage, and age at diagnosis and, thus, have independent prognostic value (appendix 2 p 15).

We thus developed an accessible WGS risk classification framework to enhance ER-positive, HER2-negative prognostication for clinical utility. WGS features that confer prognostic value independently of each other and independently of existing clinical metrics (figure 3B; appendix 1 p 28), specifically *TP53* driver mutation status, SV burden, and SBS2 and SBS13 activity, were modelled with a stepwise approach. The dataset was first split by the largest effect factor (*TP53* status). The resultant groups were then tested to see if splitting by the next factor (SV burden) could prognosticate followed by the last factor (SBS2 and SBS13 percentage; appendix 1 p 28). Our investigation yielded two groups at high risk of poor survival: a group with wild-type *TP53*, high SV, and high SBS2 and SBS13 and a group with *TP53* mutation (without high SV or high SBS2 and SBS13). There were two groups with better outcomes: a group at low-to-medium risk with wild-type *TP53*, high SV burden but

low SBS2 and SBS13, and a group at low risk comprising tumours with wild-type *TP53* and low SV burden (low-risk *vs* high-risk mutated *TP53*: HR 7·6 [95% CI 4·0–14·0; p<0·0001; low-risk *vs* high-risk wildtype *TP53*: 6·8 [3·1–15·0]; p<0·0001; and low-risk *vs* low-to-medium risk: 3·1 [1·5–6·4]; p=0·0018; appendix 2 p 15). Notably, although high-risk genomic groups were enriched for high-grade and late-stage cancers, we also observed a substantial proportion of grade 1 or 2 tumours (132 [44·3%] of 298) and stage I or II tumours (150 [50·3%]). These genomic risk groups remained significantly prognostic when grade 2 and grade 3 cancers were analysed separately (figure 3C; appendix 1 p 29). Given that both groups at high risk conferred similar survival associations, they were combined in further analyses.

In the validation of our framework, the SCAN-B cohort recapitulated our results between the groups at low risk versus high risk for distant recurrence-free interval (HR 2·4 [95% CI 1·3–4·5; p=0·0046) and overall survival (1·6 [1·1–2·5]; p=0·028; figure 3D; appendix 2 pp 22–25). We further assessed our framework's performance correcting for grade, age, and treatment group in the SCAN-B cohort. The prognosticator remained significant between the groups at low risk versus high risk for distant recurrence-free interval (2·1 [1·0–4·2]; p=0·048) and overall survival (1·7 [1·0–2·9]; p=0·035; appendix 2 p 25).

When comparing the prognosticator with multigene expression predictors favoured in high-income countries and institutions, the WGS-based framework was the only prognosticator that significantly predicted overall survival, after correcting for age, grade, and treatment group (appendix 1 p 30; appendix 2 p 25). This observation is likely because the prognosticator offers independent WGS-based information to augment current clinical prognosticators, compared with multigene classifiers that incorporate clinical factors (eg, ER, PR, Ki67, tumour size, nodal status, and proliferation) as part of their calculation.

Discussion

This retrospective analysis is the largest population-based WGS study on breast cancer integrated with national data on cancer-specific mortality. First, we report highly personalised genomic information in 27% of breast cancers, whether for immediate actionability or for prospective clinical trials, equating to bringing potential clinical impact to more than 15 000 patients with breast cancer per year in the UK. Our findings suggest an enrichment of potentially under-diagnosed, inadequately treated tumours with HRD in the ER-positive, HER2-negative breast cancer subtype that progress to metastatic disease. These observations argue for a clinical trial to investigate the effects of compounds, such as PARP inhibitors in this subset of patients, beyond germline *BRCA1* or *BRCA2* carriers. For example, a phase 2 window trial of the PARP inhibitor rucaparib in patients with treatment-naive triple-negative breast cancer showed activity in patients with HRD, regardless of the cause of HRD (eg, germline, somatic, or promoter hypermethylation; EudraCT2014-003319-12;[34] appendix 2 p 20). If effective in ER-positive, HER2-negative breast cancer, a similar opportunity could translate to affecting approximately 2900 patients per year in the UK. Targeted sequencing approaches are estimated to miss 1200 of these patients annually, including cases that are *BRCA1* or *BRCA2* wild-type, promoter hypermethylated, or affected by SVs (either germline or somatic).

Second, we present a prognosticator to facilitate future clinical studies. By use of the WGS risk framework, approximately 7500 new diagnoses of breast cancer with low-grade but genomically high-risk tumours would be identifiable per year in the UK. This prognostication allows for increased intervention; for example, CDK4/6 inhibitors and extended hormonal therapy for patients without personalised features. Conversely, more than 22 000 patients per year in the UK with clinically and genomically low-risk tumours could be eligible for treatment de-escalation with appropriate monitoring.

Although the 100kGP breast cancer cohort has the advantage of being representative of the UK population, it also has inbuilt limitations. First, we were unable to analyse the effect of treatment on outcome due to heterogeneity in clinical practice and irregular annotation. Second, because transcriptomic assays are not routinely used in low-resourced, non-tertiary sites in England, these data were unavailable; the most consistent clinical classification approach is immunohistochemistry and grade. To address these limitations, we used data from additional consortia, predominantly the SCAN-B cohort, which benefits from high-quality data on treatment, histology, and outcomes.

WGS offers holistic, comprehensive genomic reporting for each patient, showing all driver mutations, mutational signatures, and prognostic biomarkers in a single assessment. These discoveries come at a time when logistical challenges to the widespread implementation of WGS in the clinic have diminished; cold storage requirements have now been eliminated, and RNAlater-preserved tissue produces WGS data indistinguishable from fresh–frozen data.[17] Additionally, formalin fixation and paraffin-embedded material are usable alternatives when necessary.[35] To truly revolutionise genomics in cancer care, we propose that WGS should be used as a triaging step, with data considered in two stages. First, use the data to seek highly personalised and biologically distinguishing features for clinical intervention or clinical trials. Then, in the absence of such features, use the data to inform prognosis or the most appropriate clinical strategy or trial going forward. Several aspects of this proposition are already routinely practised. For example, if a tumour were HER2-positive or ER-positive, the use of anti-HER2 strategies or endocrine therapy is

continued. However, if at triage, a tumour carried a personalised feature (eg, HRD, mismatch repair deficiency, or other rare features), then selective therapeutics could be considered as part of a prospective clinical trial. For tumours without individualised features, our WGS-based risk framework could distinguish between individuals at low risk, for whom potential therapy de-escalation would be a possibility, and those at high risk, for whom alternative strategies would need to be considered. In summary, this work calls for a shift in mindset regarding the use of WGS as a holistic readout to gather the necessary evidence base to support true genome-directed precision medicine.

For **Genomics England** see https://www.genomicsengland.co.uk/research

Contributors

SNZ, DB, HRD, and NT conceptualised the study. DB, SNZ, GCCK, HRD, AD, YM, JC, YK, LC, SJZ, JW, GR, and SS contributed to the methodology. MC, GCC, PA, AS, JA, and MAB curated the data. DB, HRD, MC, GCC, AD, LC, and JS accessed and verified the data. DB, SNZ, HRD, GCCK, and LC contributed to the investigation. DB, CB, NT, and JS conducted the formal analysis. DB and SNZ produced the figures. MAB, AS, JA, PA, MCD, VD, RD, and AB contributed resources. SN-Z acquired the funding. SN-Z, NT, EC, and CP contributed to project administration. SNZ and HD supervised the study. DB and SNZ wrote the original draft of the manuscript. SNZ, DB, NT, JS, HRD, GCCK, JC, PJT, and SB wrote, reviewed, and edited the manuscript.

Declaration of interests

HRD, GR, GCCK, and SN-Z are inventors on patent applications encompassing the code and intellectual principle of the HRDetect algorithm (PCT/EP2017/060294; HRD and SN-Z), clinical use of signatures (PCT/EP2017/060289; HRD and SN-Z), clinical predictors (PCT/EP2017/060298; HRD and SN-Z), rearrangement signature methods (PCT/EP2017/060279; SNZ), indel signature methods (PCT/EP2024/077959; GCCK and SN-Z), and PRRDetect algorithm (PCT/EP2024/078030; GCCK, GR, and SN-Z). MAB reports consulting fees from Atlis Therapeutics and Grey Wolf Therapeutics; payments for expert testimony from Clementia and Ipsen; participation on a data safety monitoring board or advisory board for Clementia, Ipsen, and Incyte; and grants or contracts from UCB not supporting the present work. CP has received consulting fees from Pfizer, AstraZeneca, Daiichi Sankyo, Eli Lilly, Eisai, Exact Sciences, Gilead, Medac, MSD, Novartis, Roche, and Seagen; payments for honoraria lectures from Pfizer, AstraZeneca, Seagen, and Novartis; and support for attending meetings or travel from Gilead, Roche, and Novartis. CP reports a leadership or fiduciary role in other board, society, committee, or advocacy group (paid or unpaid) with National Cancer Research Institute; receipt of equipment, materials, drugs, medical writing, gifts, or other services from Pfizer and Seagen; and grants or contracts not supporting the present work from Daiichi Sankyo, Pfizer, Gilead, and Seagen. CP is also an advisor for Make Seconds Count. EC reports consulting fees from Pfizer; payments for honoraria lectures from Pfizer, AstraZeneca, Novartis, Eli Lilly, Menarini Stemline, Roche, and Guardant; support for attending meetings or travel from Roche; and a leadership or fiduciary role in other board, society, committee, or advocacy group (paid or unpaid) with Chair of Expert Committee on Survivorship. EC also reports receipt of equipment, materials, drugs, medical writing, gifts, or other services from Seca; participation on a data safety monitoring board or advisory board for AstraZeneca, Guardant, Eli Lilly, Novartis, Pfizer, and Roche; grants or contracts not supporting the present work from Daiichi Sankyo and AstraZeneca; and a research collaboration with Proteotype. AB reports payments for honoraria lectures from AstraZeneca. SNZ reports support for attending meetings or travel from the American Association of Cancer Research, the European Association of Cancer Research, European Society of Human Genetics, the British Society for Genetic Medicine, Fusion conferences, Gordon Research Conference, and European Molecular Biology Organization conferences; and participation on the UK MHRA Committee for Highly Personalised Therapies advisory board. NT reports grants or contracts not supporting the present work from Pfizer, AstraZeneca, Roche–Genentech, Merck Sharpe and Dohme, Guardant Health, Invitae, Invitae Personalis, and Natera; receipt of advisory board honoraria from Astra Zeneca, Lilly, Pfizer, Roche–Genentech, Novartis, GlaxoSmithKline, Repare Therapeutic, Invitia, Guarda Health, and Exact sciences; and funding from AstraZeneca, Pfizer, Roche–Genentech, Merck Sharpe and Dohme, Guardant Health, Invitae, Inivata, Personalis, and Natera. PA is the non-executive director of Health and Data Research UK and has stock options with Amazon.

Data sharing

Research on the de-identified patient data used in this Article can be carried out in the Genomics England Research Environment, subject to a collaborative agreement that adheres to patient-led governance. All interested readers will be able to access the data in the same manner that the authors accessed the data. For more information about accessing the data, interested readers can contact research-network@genomicsengland.co.uk or access the relevant information on the Genomics England website.

Acknowledgments

This work was funded by the National Institute of Health Research (NIHR) Professorship (NIHR301627; SN-Z), the Breast Cancer Research Foundation (BCRF-24-197; SN-Z), the Dr Josef Steiner Cancer Research Award 2019 (SN-Z), the Basser Gray Prime Award 2020 (SN-Z), Cancer Research UK (Grand Challenge Award C60100/A25274 and CGCATF-2021/100013; Advanced Clinician Scientist Award C60100/A23916; SN-Z), and and this research was supported by the NIHR Cambridge Biomedical Research Centre (NIHR203312; BRC-1215-20014). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. GCCK was funded by the Sir Jeffrey Cheah Early Career Fellowship. Analysis conducted on the validation SCAN-B cohort was funded by the Swedish Cancer Society (CAN 2024/3591 22 2143 Pj 01 H; JS and AB), the Swedish Research Council (2021-01800; JS), the Fru Berta Kamprads Foundation (FBKS-2023-21; AB), and the Mats Paulsson Foundation (23500 Lund-Lausanne; AB). This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England (a wholly owned company of the UK Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care, as well as data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK, and the Medical Research Council have also funded research infrastructure. We thank the large network of individuals who contributed to the breast Genomics England Clinical Interpretation Partnership for their efforts, particularly with regard to patient recruitment and sample acquisition, without which this research would not be possible. We also acknowledge patients and clinicians participating in the SCAN-B study, personnel at the central SCAN-B laboratory at the Division of Oncology, Lund University; the Swedish National Breast Cancer Quality Registry; Regional Cancer Center South; Regional Biobank Center South; and the South Sweden Breast Cancer Group.

References

1 Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; **74:** 229–63.

2 Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA* 2019; **321:** 288–300.

3 Day S, Coombes RC, McGrath-Lone L, Schoenborn C, Ward H. Stratified, precision or personalised medicine? Cancer services in the 'real world' of a London hospital. *Sociol Health Illn* 2017; **39:** 143–58.

4 Tsimberidou AM, Sireci A, Dumanois R, Pritchard D. Strategies to address the clinical Practice gaps affecting the implementation of personalized medicine in cancer care. *JCO Oncol Pract* 2024; **20:** 761–66.

5 Schrag D, Kuntz KM, Garber JE, Weeks JC. Life expectancy gains from cancer prevention strategies for women with breast cancer and *BRCA1* or *BRCA2* mutations. *JAMA* 2000; **283:** 617–24.

6 Zardavas D, Te Marvelde L, Milne RL, et al. Tumor *PIK3CA* genotype and prognosis in early-stage breast cancer: a pooled analysis of individual patient data. *J Clin Oncol* 2018; **36:** 981–90.

7 Nik-Zainal S, Alexandrov LB, Wedge DC, et al, and the Breast Cancer Working Group of the International Cancer Genome Consortium. Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012; **149:** 979–93.

8 Degasperi A, Zou X, Amarante TD, et al, and the Genomics England Research Consortium. Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* 2022; **376:** abl9283.

9 Bertucci F, Ng CKY, Patsouris A, et al. Genomic characterization of metastatic breast cancers. *Nature* 2019; **569:** 560–64.

10 Koh GCC, Nanda AS, Rinaldi G, et al. A redefined InDel taxonomy provides insights into mutational signatures. *Nat Genet* 2025; **57:** 1132–41.

11 Angus L, Smid M, Wilting SM, et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet* 2019; **51:** 1450–58.

12 Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016; **534:** 47–54.

13 Davies HR, Black D, Kvist A, et al. Homologous recombination deficiency in primary ER-positive and *HER2*-negative breast cancer. *medRxiv* 2025; published online June 4. https://doi.org/10.1101/2025.06.04.25328880.

14 Caulfield M, Davies J, Dennys M, et al. The National Genomics Research Library v5.1: amendment to The National Genomics Research and Healthcare Knowledgebase v5. Genomics England Limited, 2018.

15 Davies H, Glodzik D, Morganella S, et al. HRDetect is a predictor of *BRCA1* and *BRCA2* deficiency based on mutational signatures. *Nat Med* 2017; **23:** 517–25.

16 Fumet JD, Truntzer C, Yarchoan M, Ghiringhelli F. Tumour mutational burden as a biomarker for immunotherapy: Current data and emerging concepts. *Eur J Cancer* 2020; **131:** 40–50.

17 Staaf J, Glodzik D, Bosch A, et al. Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. *Nat Med* 2019; **25:** 1526–33.

18 Schwartz CJ, da Silva EM, Marra A, et al. Morphologic and genomic characteristics of breast cancers occurring in individuals with Lynch syndrome. *Clin Cancer Res* 2022; **28:** 404–13.

19 Barroso-Sousa R, Pacífico JP, Sammons S, Tolaney SM. Tumor mutational burden in breast cancer: current evidence, challenges, and opportunities. *Cancers (Basel)* 2023; **15:** 3997.

20 van't Veer LJ, Paik S, Hayes DF. Gene expression profiling of breast cancer: a new tumor marker. *J Clin Oncol* 2005; **23:** 1631–35.

21 Cheung S, Greenway N, Lagord C, et al. All breast cancer report: a UK analysis of all symptomatic and screen-detected breast cancers diagnosed in 2006. https://associationofbreastsurgery.org.uk/media/aecde1go/all-breast-cancer-report1.pdf (Sept 3, 2025).

22 Stephens PJ, Tarpey PS, Davies H, et al, and the Oslo Breast Cancer Consortium (OSBREAC). The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012; **486:** 400–04.

23 Kleivi K, Lothe RA, Heim S, et al. Genome profiling of breast cancer cells selected against in vitro shows copy number changes. *Genes Chromosomes Cancer* 2002; **33:** 304–09.

24 Wagner VP, Bingle CD, Bingle L. *MYB-NFIB* fusion transcript in adenoid cystic carcinoma: Current state of knowledge and future directions. *Crit Rev Oncol Hematol* 2022; **176:** 103745.

25 Vocka M, Zimovjanova M, Bielcikova Z, et al. Estrogen receptor status oppositely modifies breast cancer prognosis in *BRCA1/BRCA2* mutation carriers versus non-carriers. *Cancers (Basel)* 2019; **11:** 738.

26 Reijns MAM, Parry DA, Williams TC, et al, and the Genomics England Research Consortium, and the Colorectal Cancer Domain UK 100,000 Genomes Project. Signatures of *TOP1* transcription-associated mutagenesis in cancer and germline. *Nature* 2022; **602:** 623–31.

27 Horton JK, Stefanick DF, Prasad R, Gassman NR, Kedar PS, Wilson SH. Base excision repair defects invoke hypersensitivity to *PARP* inhibition. *Mol Cancer Res* 2014; **12:** 1128–39.

28 Jeselsohn R, Buchwalter G, De Angelis C, Brown M, Schiff R. *ESR1* mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat Rev Clin Oncol* 2015; **12:** 573–83.

29 Razavi P, Chang MT, Xu G, et al. The genomic landscape of endocrine-resistant advanced breast cancers. *Cancer Cell* 2018; **34:** 427–438.

30 Cancer Research UK. Breast cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer (accessed Sept 3, 2025).

31 Bailey C, Pich O, Thol K, et al, and the Genomics England Consortium. Origins and impact of extrachromosomal DNA. *Nature* 2024; **635:** 193–200.

32 Houlahan KE, Mangiante L, Sotomayor-Vivas C, et al. Complex rearrangements fuel ER^+ and $HER2^+$ breast tumours. *Nature* 2025; **638:** 510–18.

33 Mollon LE, Anderson EJ, Dean JL, et al. A systematic literature review of the prognostic and predictive value of PIK3CA mutations in $HR^+/HER2^-$ metastatic breast cancer. *Clin Breast Cancer* 2020; **20:** e232–43.

34 Chopra N, Tovey H, Pearson A, et al. Homologous recombination DNA repair deficiency and PARP inhibition activity in primary triple negative breast cancer. *Nat Commun* 2020; **11:** 2662.

35 Basyuni S, Heskin L, Degasperi A, et al, and the PARTNER Trial Group, and the Personalised Breast Cancer Program Group. Large-scale analysis of whole genome sequencing data from formalin-fixed paraffin-embedded cancer specimens demonstrates preservation of clinical utility. *Nat Commun* 2024; **15:** 7731.